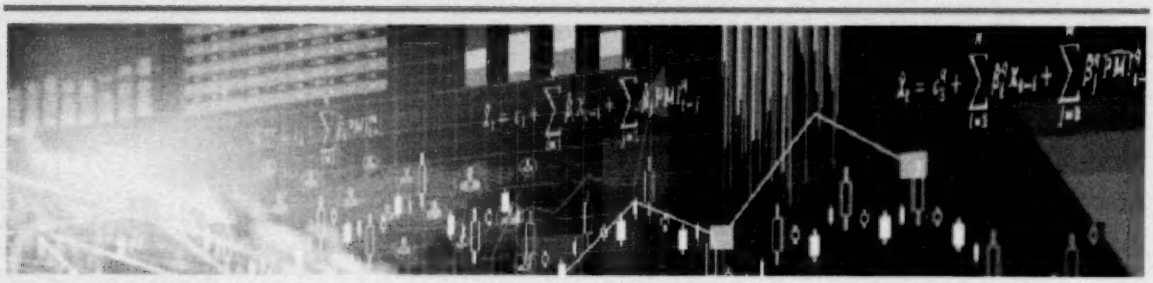
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2016-22

Estimating Systematic Risk Under Extremely Adverse Market Conditions



by Maarten R. C. van Oordt and Chen Zhou

Bank of Canada staff working papers provide a forum for staff to publish work-in-progress research independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank of Canada or De Nederlandsche Bank.

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2016-22

May 2016

Estimating Systematic Risk Under Extremely Adverse Market Conditions

by

Maarten R. C. van Oordt[1] and Chen Zhou[2]

[1]Financial Stability Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
mvanoordt@bankofcanada.ca

[2]Economics and Research Division
De Nederlandsche Bank
1000AB Amsterdam, The Netherlands
and
Erasmus University Rotterdam
C.Zhou@DNB.nl
zhou@ese.eur.nl

ISSN 1701-9397

© 2016 Bank of Canada

Acknowledgements

The authors thank participants in various conferences and seminars where previous versions of this paper have been presented: the Spring Meeting of Young Economists (Groningen, 2011), the workshop on Extreme Dependence in Financial Markets (Rotterdam, 2011), the 9th INFINITI Conference on International Finance (Dublin, 2011), the 7th Conference on Extreme Value Analysis (Lyon, 2011), the 65th Econometric Society European Meeting (Oslo, 2011), the Joint Statistical Meetings (Miami, 2011), the 8th International Conference of the ERCIM WG on Computational and Methodological Statistics (London, 2015), and seminars in De Nederlandsche Bank, the Ca' Foscari University of Venice, the Federal Reserve Bank of Cleveland and the University of Amsterdam. Especially, the authors want to thank Jason Allen, Kris M. R. Boudt (discussant), Bruno Feunou, Bernd Schwaab (discussant) and Casper G. de Vries. An earlier version of this paper was distributed under the title "Systematic risk under extremely adverse market conditions."

Abstract

This paper considers the problem of estimating a linear model between two heavy-tailed variables if the explanatory variable has an extremely low (or high) value. We propose an estimator for the model coefficient by exploiting the tail dependence between the two variables and prove its asymptotic properties. Simulations show that our estimation method yields a lower mean squared error than regressions conditional on tail observations. In an empirical application we illustrate the better performance of our approach relative to the conditional regression approach in projecting the losses of industry-specific stock portfolios in the event of a market crash.

JEL classification: C14, G01
Bank classification: Econometric and statistical methods; Financial markets

Résumé

Dans cet article, nous examinons l'estimation d'un modèle de régression linéaire entre deux variables qui suivent des lois de probabilité à queue épaisse lorsque la variable explicative a des valeurs extrêmement faibles (ou élevées). En exploitant la relation de dépendance entre les extrema des deux variables, nous proposons un estimateur du coefficient de régression. Nous établissons les propriétés asymptotiques de cet estimateur. Les simulations montrent que notre méthode d'estimation génère une plus faible erreur quadratique moyenne que l'outil de référence, c'est-à-dire l'estimateur des moindres carrés ordinaires obtenu à partir des observations extrêmes. Nous illustrons également la supériorité de notre approche par rapport à l'outil de référence dans la prévision des pertes qu'enregistreraient des portefeuilles d'actions propres à un secteur particulier en cas de crash boursier.

Classification JEL : C14, G01
Classification de la Banque : Méthodes économétriques et statistiques; Marchés financiers

NON-TECHNICAL SUMMARY

For many decision makers, it is important to analyze scenarios considering extremely adverse market conditions. For example, investors are concerned with the performance of stock portfolios in a market crash, while policy-makers are concerned with the performance of financial institutions in systemic events. In this paper, we develop a new method for evaluating the consequences of such extreme events.

A difficulty with assessing potential outcomes in market crashes is that interrelationships in financial markets tend to change in extremely adverse market conditions. For example, correlations in financial markets tend to be stronger in a market crash than in market booms. Moreover, stock market returns are governed by heavy tails and exhibit tail dependence. Neglecting such issues may result in serious flaws in projections for such stress scenarios. Therefore, it is important to apply a method that takes these aspects into account.

The method developed in this paper estimates the sensitivity of firms' stock returns to extremely adverse shocks in the market. The estimation method relies on Extreme Value Theory and assesses the co-movement based on large shocks in the past. Because observations with large shocks are sparse in the data, an efficient use of those observations is of utmost importance. We show via simulations that the developed method performs relatively well if the estimation is based on relatively few so-called "tail events." Moreover, in an empirical application, we assess its historical performance in projecting the losses of industry-specific stock portfolios in major market crashes over the past 80 years.

1 INTRODUCTION

In financial management, the risk of stock portfolios is often assessed by estimating their return sensitivity to key risk factors. The coefficient in a single-factor model, the "market beta" is commonly given a prominent position in such an assessment. Nevertheless, there is wide consensus that the relationship between asset returns and market risk depends on market conditions. For example, equity returns exhibit stronger correlation during volatile periods, especially in the case of extreme market downturns; see, e.g., King and Wadhwani (1990), Longin and Solnik (1995, 2001) and Ang and Chen (2002). Thus, risk managers who are concerned about possible extreme losses in distress events may need to analyze systematic risk only under extremely adverse market conditions. In this paper, we develop a new method for evaluating the consequences of such extreme events.

The goal of this paper is to estimate a linear model between two heavy-tailed variables conditional on the explanatory variable having an extremely low (or high) value. Consider the following model on the relation between two random variables (rvs) X and Y, conditional upon an extremely low value of X

$$Y = \beta^T X + \varepsilon, \quad \text{for} \quad X < Q_x(p), \tag{1.1}$$

where p denotes a very small probability, ε is the error term that is assumed to be independent of rv X under the condition $X < Q_x(p)$, and where $Q_x(p)$ denotes the quantile of rv X at the probability level p. We intend to estimate coefficient β^T using observations on (X, Y). With Y and X as the returns of, respectively, a stock portfolio and the market portfolio, the coefficient β^T is regarded as a measure of systematic risk under extremely adverse market conditions. Estimating β^T can be useful to assess the extreme loss on the stock portfolio in the event of a market crash.

The coefficient β^T in the linear tail model in Eq. (1.1) can be regarded as a regression coefficient. Consequently, a direct approach to estimating β^T is to apply a "conditional

regression," i.e., to estimate a least squares regression coefficient based on observations corresponding to extremely low values of X only. This method has been used by, for example, Mitchell and Pulvino (2001) to evaluate the β^T of a trading strategy based on "merger arbitrage" and by Post and Versijp (2007) to estimate the β^Ts of low beta stocks.

Two potential drawbacks of this conditional regression approach apply. First, because the conditional regression is based on a small number of observations, the approach may potentially produce a relatively large variance of the estimator. Second, when applied to financial market data, the heavy-tailedness of financial returns may further increase the estimation error.

We propose an alternative estimator for β^T by exploiting the tail dependence imposed by the heavy-tailedness of X and Y. Under mild conditions, we show that the proposed estimator possesses consistency and asymptotic normality. Simulations show that our estimation method yields a lower mean squared error than estimating conditional regressions on tail observations. In an empirical application, we illustrate the better performance of our approach relative to the conditional regression approach in projecting the losses of industry-specific portfolios in major stock market crashes over the past 80 years.

Theoretically, our estimator has a structure similar to a regression coefficient. The estimator in a standard univariate regression analysis consists of a dependence measure given by the correlation, and the marginal risk measures given by the standard deviations. In the estimator of β^T, the dependence measure is replaced by a tail dependence measure and the marginal risk measures are replaced by quantiles obtained from tail observations.

Our study builds on the literature on multivariate Extreme Value Theory (EVT) that provides measures to evaluate the dependence among extreme observations; see, e.g., Embrechts et al. (2000) and De Haan and Ferreira (2006, Chapters 6 and 7). Poon et al. (2004) and Hartmann et al. (2004) have applied such tail dependence measures to analyze linkages between asset returns during crises. Bollerslev et al. (2013) analyze the tail dependence between jumps in stock returns using high-frequency data. Malevergne and Sornette (2004)

derive the level of tail dependence, assuming an unconditional linear relation with a single risk factor. Others study tail dependence assuming global linear relations with multiple risk factors; see, e.g., De Vries (2005) and Hartmann et al. (2010). In contrast, our approach focuses on estimating β^T by exploiting the tail dependence structure while assuming the linear model only in the tail.

Our study should be distinguished from the literature on estimating unconditional regression models in the presence of heavy tails. Mikosch and De Vries (2013) show that the finite sample distribution of a regression coefficient is heavy-tailed if the error term follows a heavy-tailed distribution. This may be improved by applying least (tail-)trimmed squares, which ensures asymptotic normality even if the error terms have infinite variance; see Rousseeuw (1985) and Hill (2013). Our study differs from this existing literature in the sense that our purpose is to estimate the linear relation only for extremely low X, rather than estimating an unconditional linear relation. Similarly, our study is distinct from quantile regressions; see, e.g., Koenker and Bassett (1978). Quantile regression analysis focuses on predicting the quantile of Y using an unconditional relation to X for all potential values of X. We focus on predicting the expectation of Y using a linear relation with X conditional on an extreme value of X. Finally, our study should also be differentiated from models in which the dependent variable reacts differently to diffusive and jump components in the independent variable; see Todorov and Bollerslev (2010). The linear coefficient on the jump component is different from β^T. Given an extremely low realization of the independent variable, this realization could potentially be attributed mainly to the jump component but, nevertheless, also involves a diffusive component. Therefore, β^T is a combination of the two linear coefficients on the diffusive and jump components, albeit mainly loaded on the latter. Different from Todorov and Bollerslev (2010), the linear tail model in (1.1) does not require a symmetric relation to extremely low or high values of the independent variable: By applying our methodology to $-Y$ and $-X$, one obtains an estimate of the (potentially different) relation when X has an extremely high value.

The remainder of the paper is organized as follows. Section 2 describes our estimation method and reports several simulation results. Section 3 provides an empirical application of our approach in which we estimate the losses on industry-specific stock portfolios during market crashes. Section 4 concludes.

2 METHODOLOGY

2.1 Theory

We start by assuming heavy-tailedness of X and Y. The definition of heavy-tailedness is as follows. The tail distributions of X and Y are heavy-tailed if they can be expressed as

$$\Pr(X < -u) \sim u^{-\alpha_x} l_x(u) \text{ and } \Pr(Y < -u) \sim u^{-\alpha_y} l_y(u), \quad \text{as } u \to \infty, \tag{2.1}$$

where l_x and l_y are slowly varying functions as $u \to \infty$. That is,

$$\lim_{u\to\infty} \frac{l_x(tu)}{l_x(u)} = \lim_{u\to\infty} \frac{l_y(tu)}{l_y(u)} = 1,$$

for any $t > 0$. Parameters α_x and α_y are called the tail indices.

The idea behind our approach to estimating β^T is as follows. The relation in Eq. (1.1) is specified only for the region $X < Q_x(\bar{p})$, while the model specifies no assumptions on the relation for the region $X \geq Q_x(\bar{p})$. The relation brings about a dependence structure between X and Y in the case of extremely low values of X, i.e., if $X < Q_x(\bar{p})$. This structure determines the dependence between the left tails of the distributions of X and Y. Our approach relies on analyzing this tail dependence structure to infer the level of β^T.

We consider the following tail dependence measure from multivariate EVT,

$$\tau := \lim_{p\to 0} \tau(p) := \lim_{p\to 0} \frac{1}{p} \Pr(Y < Q_y(p), X < Q_x(p)), \tag{2.2}$$

where $Q_x(p)$ and $Q_y(p)$ are the quantiles of X and Y at probability level p.[1] The tail dependence measure can be rewritten as $\tau = \lim_{p \to 0} \Pr(Y < Q_y(p) | X < Q_x(p))$, which is the probability of observing an extremely low value of Y conditional on an extremely low value of X. Since it is the limit of a conditional probability, the τ-measure is bounded by $0 \leq \tau \leq 1$. The case $\tau = 0$ is regarded as tail independence, while the case $\tau = 1$ corresponds to complete tail dependence. Also, the tail dependence measure is invariant to positive linear transformations on X and Y. These features of the τ-measure indicate that its role in our approach will resemble that of a correlation coefficient, except that the τ-measure focuses on dependence in the tails only.

The following theorem shows how the τ-measure relates to the coefficient β^T in the linear tail model in Eq. (1.1).

Theorem 1 *Under the linear tail model in Eq.* (1.1) *and the heavy-tail set-up of the down-side distributions in* (2.1)*, with* $\alpha_y > \frac{1}{2}\alpha_x$ *and* $\beta^T \geq 0$*, we have that*

$$\lim_{p \to 0} (\tau(p))^{1/\alpha_x} \frac{Q_y(p)}{Q_x(p)} = \beta^T. \tag{2.3}$$

Proof. See the Appendix. ■

Theorem 1 does not depend on assuming the heavy-tailedness of the unobservable error term ε. The theorem holds even if ε exhibits a thin-tailed distribution, such as the normal distribution. The condition $\alpha_y > \frac{1}{2}\alpha_x$ basically requires Y not to be "too heavy-tailed" in comparison with X. The intuition is that, otherwise, the error terms ε would have a much heavier tail than X. The impact of extreme realizations of ε on extreme realizations of Y would overshadow the impact of the relation between X and Y. As a consequence, it is not possible to infer the level of β^T. Nevertheless, this condition is not very restrictive in the context of stock market returns. For example, if X represents the returns on a general market index with an α_x of 4 (see, e.g., Jansen and De Vries (1991)), the condition is satisfied

if the firm's stock returns Y have finite variance. Moreover, conditional upon a sufficiently low α_x, Theorem 1 also holds if Y has infinite variance or mean.

The relation in Theorem 1 provides the basis for the estimation of coefficient β^T. Consider independent and identically distributed (i.i.d.) observations $(X_1, Y_1), \cdots, (X_n, Y_n)$ with the i.i.d. unobserved error terms $\varepsilon_1, \cdots, \varepsilon_n$. Later we will also consider the presence of temporal dependence. To estimate β^T, we estimate each component in Eq. (2.3). As in usual extreme value analysis, we mimic the limit procedure $p \to 0$ by considering only the lowest k observations in the tail region, such that $k := k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. In other words, for statistical estimation, the probability p is set at some low level $p = k/n$. Hence, we obtain the estimator of β^T as

$$\hat{\beta}^T := \hat{\tau}(k/n)^{1/\hat{\alpha}_x} \frac{\hat{Q}_y(k/n)}{\hat{Q}_x(k/n)}. \tag{2.4}$$

We remark that the estimator $\hat{\beta}^T$ in Eq. (2.4) shows similarities with a standard regression analysis. Considering a standard linear regression between random variables U and V, the estimator of the slope coefficient is $\hat{\rho}\hat{\sigma}_u/\hat{\sigma}_v$, where $\hat{\rho}$ is the correlation coefficient between U and V, and where $\hat{\sigma}_u$ and $\hat{\sigma}_v$ are the standard deviations of U and V, respectively. Similarly, the estimator $\hat{\beta}^T$ consists of the tail dependence measure $\hat{\tau}$, and two tail risk measures, i.e., the tail quantiles of X and Y. In addition, it combines these components in a similar way as in a standard regression analysis.

2.2 Estimation

For our procedure, we rely on relatively simple and widely used estimators to obtain estimates of each of the components in Eq. (2.4). These estimators rely exclusively on observations far in the tail of the distributions of X and Y. Nevertheless, the development of better estimators for the building blocks in Eq. (2.4) has been the subject of an extant literature. Hence, our procedure to estimate β^T via Eq. (2.4) may well stand to be further improved

by choosing other estimators for the components. Throughout this paper, we will refer to our estimation of β^T with the estimators of the components below as the EVT approach.

The estimate of the tail index α_x is obtained from the k_1 lowest observations of X with the estimator proposed in Hill (1975). Here, k_1 is another intermediate sequence such that $k_1 := k_1(n) \to \infty$ and $k_1/n \to 0$ as $n \to \infty$. Suppose the observations of (X, Y) are $(X_1, Y_1), \cdots, (X_n, Y_n)$. By ranking the observations of X_i as $X_{n,1} \leq X_{n,2} \leq \cdots \leq X_{n,n}$, the Hill estimator is defined as

$$\frac{1}{\hat{\alpha}_x} := \frac{1}{k_1} \sum_{i=1}^{k_1} \log \left(\frac{X_{n,i}}{X_{n,k_1+1}} \right). \tag{2.5}$$

For the τ-measure, multivariate EVT provides a nonparametric estimate; see Embrechts et al. (2000). The estimator is given as

$$\hat{\tau}(k/n) := \frac{1}{k} \sum_{i=1}^{n} \mathbf{1}_{\{Y_i < Y_{n,k+1}, X_i < X_{n,k+1}\}}, \tag{2.6}$$

where $Y_{n,k+1}$ is the $(k+1)$-th lowest order statistic of Y_i. Finally, the quantiles of X and Y at the probability level k/n, the $\hat{Q}_x(k/n)$ and $\hat{Q}_y(k/n)$, are estimated by their $(k+1)$-th lowest order statistics, i.e., $X_{n,k+1}$ and $Y_{n,k+1}$. Notice that, in Eq. (2.3), the same tail probability p appears in the term $\tau(p)$ and the two quantiles $Q_y(p)$ and $Q_x(p)$. Correspondingly for the estimators of the τ-measure and the quantiles of X and Y, the same intermediate sequence k is used. Differently, there is no theoretical restriction such that $k = k_1$, though this will be the most complicated case when dealing with the asymptotic normality below.

In general, the estimator of β^T via Eq. (2.4) inherits its consistency and asymptotic normality from the consistency and asymptotic normality of the estimators of its subcomponents. In addition, the estimator of β^T is even consistent if $\lim_{p \to 0} \tau(p) = 0$, even though the statistical properties of the estimator of the τ-measure are less known in this case. To prove the consistency of $\hat{\beta}^T$ also for the case $\lim_{p \to 0} \tau(p) = 0$, we require some additional conditions to ensure the asymptotic normality of the Hill estimator. These conditions are as

follows. First, the tail distribution of X follows the usual second order condition (see, e.g, De Haan and Stadtmüller (1996)) as

$$\lim_{u\to\infty} \frac{\frac{\Pr(X<-ux)}{\Pr(X<-u)} - x^{-\alpha_x}}{\eta(u)} = x^{-\alpha_x} \frac{x^{-\gamma}-1}{-\gamma}, \tag{2.7}$$

with an eventually positive or negative function $\eta(t)$ and $\gamma \geq 0$. Second, we require a condition on k_1 ensuring that k_1 is not too high:

$$\lim_{n\to\infty} \sqrt{k_1}\eta\left(-Q_x\left(\frac{k_1}{n}\right)\right) = \lambda \text{ finite.} \tag{2.8}$$

These two conditions are usually assumed to obtain the asymptotic normality of the Hill estimator; see, e.g., De Haan and Ferreira (2006), conditions (3.2.5) and (3.2.6).

Finally, an additional restriction ensures that k_1 is not too low. As $n \to \infty$,

$$\frac{\sqrt{k_1}}{\log n} \to +\infty. \tag{2.9}$$

The following theorem states the consistency of $\hat{\beta}^T$.

Theorem 2 *Suppose that* $k \to \infty$, $k_1 \to \infty$, $\frac{k}{n} \to 0$, $\frac{k_1}{n} \to 0$, *as* $n \to \infty$. *In addition, only if* $\lim_{p\to 0}\tau(p) = 0$, *do we further assume conditions* (2.7)–(2.9). *Then we have that as* $n \to \infty$,

$$\hat{\beta}^T \xrightarrow{P} \beta^T.$$

Next, we deal with asymptotic normality. For that purpose, we assume the second order condition for the distribution of Y and the joint distribution (X, Y). This is in line with usual asymptotic normality result in multivariate extreme value statistics; see, e.g., Einmahl et al. (2006). First, assume that

$$\lim_{u\to\infty} \frac{\frac{\Pr(Y<-ux)}{\Pr(Y<-u)} - x^{-\alpha_y}}{\eta'(u)} = x^{-\alpha_y} \frac{x^{-\gamma'}-1}{-\gamma'}, \tag{2.10}$$

with an eventually positive or negative function $\eta'(t)$ and $\gamma' \geq 0$.

Second, with denoting the distribution function of X and Y as F_x and F_y, we assume that as $u \to 0$

$$\frac{1}{u} \Pr\left(F_x(X) < ux, F_y(Y) < uy\right) - R(x,y) = O\left(t^{\theta}\right), \tag{2.11}$$

for some $\theta > 0$, and $(x,y) \in [0,1]^2 / \{(0,0)\}$.

The following theorem states the asymptotic normality of $\hat{\beta}^T$.

Theorem 3 *Assume that* $\lim_{p \to 0} \tau(p) = \tau > 0$ *and* $\Pr(Y > u) = O(\Pr(Y < -u))$ *as* $u \to \infty$. *Further assume that the second order conditions* (2.7),(2.10) *and* (2.11) *hold. Suppose* $k_1 = k = O(n^{\zeta})$, *where* $\zeta < \min\left(2\theta/(1+2\theta), 2\gamma/(2\gamma+\alpha_x), 2\gamma'/(2\gamma'+\alpha_y), 3/(\alpha_y+2)\right)$. *We then have that, as* $n \to \infty$,

$$\sqrt{k}\left(\hat{\beta}^T - \beta^T\right) \xrightarrow{d} N\left(0, \frac{\left(\beta^T\right)^2}{\alpha_x^2}\left(\frac{1}{\tau} - 1 - (\log\tau)^2\right)\right).$$

In this theorem we choose $k = k_1$ because this is what we use in the simulation and application. We note, however, that this choice is not necessarily the only one. One may also choose k and k_1 such that, as $n \to \infty$, k/k_1 converges to zero, a finite positive number or infinity. If $k/k_1 \to 0$ as $n \to \infty$, the asymptotic limit of $\hat{\alpha}_x$ will play a dominant role in that of $\hat{\beta}^T$. Conversely, if $k/k_1 \to \infty$ as $n \to \infty$, the asymptotic limit of $\hat{\tau}(k/n)$ and the two quantiles will play a dominant role in that of $\hat{\beta}^T$. Therefore, the asymptotic normality result turns out to be simpler in these two cases. If k/k_1 converges to a finite value other than 1, the asymptotic normality of $\hat{\beta}^T$ can be derived in a similar way, with a slightly more complicated structure for the asymptotic variance.

The consistency and asymptotic normality results are obtained when $\{(X_t, Y_t)\}$ forms an i.i.d. sample. In the context of stock returns, it is likely that $\{(X_t, Y_t)\}$ is a time series with temporal dependence. In general, under weak conditions, EVT analysis can be applied without modification to temporally dependent data; see Drees (2008) for a general discussion. More specifically, if $\{(X_t, Y_t)\}$ exhibits weak temporal dependence such as autocorrelation

or GARCH-type volatility clustering, then the consistency will not be affected. This follows from the consistency results for each component in $\hat{\beta}^T$. For the Hill estimator, see, e.g., Hsing (1991); for the quantiles and the tail dependence measure, see, e.g., Hill (2009). It is notable that temporal dependence does affect the asymptotic normality result in the sense that it may lead to a different structure of the asymptotic variance of the estimates. Therefore, in financial applications, it may be better to rely on a block bootstrap procedure to obtain adjusted standard errors.

2.3 Simulations

We run simulations to compare the performance of the proposed procedure to estimate β^T and the performance of a regression conditional on tail observations.[2] The set-up of the simulations is as follows. In each sample, we generate 1,250 random observations for X, which corresponds approximately to the length of our estimation window in the empirical exercise. The observations for Y are constructed by aggregating the simulated X and ε according to different linear models. First, we consider three "global" linear models in which the relation is unaffected by the observation of X, i.e., $\beta = \beta^T = 0.5, 1, 1.5$. Second, we consider two segmented linear models. If the observation of X is larger than the 3rd percentile of X, then the observation of Y is generated from a linear model with $\beta = 1$.[3] Otherwise, it is generated from a linear model with $\beta^T = 0.5$ and $\beta^T = 1.5$, respectively.

Several data-generating processes are considered for X and ε. The Student-t distribution is known to be heavy-tailed with the tail index equal to the degrees of freedom. We perform simulations of X and ε based on random draws from a Student-t distribution with 3, 4 and 5 degrees of freedom, which implies X and ε are heavy-tailed with tail indices of 3, 4, and 5, respectively. The choice of the parameter α is similar to the estimates in the empirical analysis. Moreover, we perform simulations where X and ε exhibit temporal dependence and are each generated from a GARCH(1,1) process, i.e., $Z_t = \sigma_{Z,t}\zeta_t$, where $\sigma^2_{Z,t} = \psi_0 + \psi_1 Z^2_{t-1} + \psi_2 \sigma^2_{Z,t-1}$, for $Z = X, \varepsilon$. The parameter choices for the simulation

with normally distributed innovations ζ_t are $(\psi_0, \psi_1, \psi_2) = (0.5, 0.11, 0.88)$, which implies X and ε are heavy-tailed with a tail index of 3.68; see Sun and Zhou (2014, Table 3). The parameter choices in another simulation based on innovations ζ_t from a standardized Student-t distribution with 8 degrees of freedom are $(\psi_0, \psi_1, \psi_2) = (0.5, 0.08, 0.91)$, which implies X and ε are heavy-tailed with a tail index of 3.82.

For each of the five models and data-generating processes, we generate 10,000 samples and estimate β^T in each sample, using both the conditional regression approach and the EVT approach. Then, by comparing the estimates with the real β^T value, we calculate the mean squared error (MSE), the estimation bias and the estimation variance for the two approaches. For brevity, we report the simulations based on the Student-t distribution with 4 degrees of freedom, because the pattern across the simulations is very similar.

The first column of Figures 1 and 2 compares the MSE between the EVT approach and the conditional regression. Under the heavy-tailed set-up, we observe a better performance with the EVT approach relative to the conditional regression, if β^T is estimated based on a few observations in the tail, i.e., for low levels of k. However, the conditional regression may perform better if more observations from the moderate level are included, i.e., for high levels of k. Nevertheless, the MSE of the EVT approach is not very sensitive to including more observations from the moderate level. The second and third columns of Figures 1 and 2 show the decomposition of the MSE into squared bias and variance. We observe that the estimates from the conditional regression bear a larger variance, while the estimation error from the EVT approach is mainly due to positive bias.[4]

3 APPLICATION

We compare the performance of the EVT approach and the conditional regression approach in an empirical application. We employ data on value-weighted returns of 48 industry-specific stock portfolios and a general market index in the US. The return series run from 1931 until

2010.[5] We divide the data into 16 five-year subperiods. We assess the performance of both approaches in projecting the losses of industry portfolios on the day of the largest market loss within each subperiod.

Within each five-year period, we estimate the coefficient β_j^T in the linear tail model with the returns on industry portfolio j as the dependent variable and the excess market returns as the right-hand-side variable. In the estimation procedure, we exclude the day on which the market portfolio suffered its largest loss in order to obtain an "out of sample" estimate for the subsequent comparison. The coefficients are estimated with both the conditional regression approach and the EVT approach. The number of observations in each subperiod is on average 1,315 and we estimate the coefficient β_j^T with $k = 25$, or, $k/n \approx 2\%$. We denote these estimates as $\hat{\beta}_{OLS,j}^T$ and $\hat{\beta}_{EVT,j}^T$, respectively.

In line with the condition $\alpha_y > \frac{1}{2}\alpha_x$ in Theorem 1, we exclude portfolios with $\hat{\alpha}_j \leq \frac{1}{2}\hat{\alpha}_m$ in each subperiod.[6] In most subperiods, no portfolios are excluded. We denote the number of portfolios excluded for this reason by S. In most subperiods, $S = 0$. Table 1 reports the average $\hat{\beta}_{EVT,j}^T$ of the remaining portfolios for each subperiod (denoted as N). Table 1 also reports the minimum and maximum $\hat{\beta}_{EVT,j}^T$ and the corresponding industry name to give an indication of the range of the $\hat{\beta}_{EVT,j}^T$s. For each subperiod, the average estimate from the EVT approach is slightly above 1. Most $\hat{\beta}_{EVT,j}^T$s fall in the range between 0.5 and 2.0. These estimates imply that most portfolios are expected to lose between half and twice as much as the market portfolio in a market crash.

Based on the β_j^T estimates, we make a projection of the losses on each portfolio j on the day that the market suffered its largest loss. For each subperiod, we report the largest loss on the market portfolio, defined as $L_m = -\min\{R_{m,1}^e, \ldots, R_{m,t}^e\}$, and the corresponding date in Table 2. We denote the actual loss on a specific industry portfolio on that day as $L_j = -R_{j,\bar{t}}^e$, where $\bar{t}$ refers to the day of the largest loss on the market portfolio. Following the linear tail model, the expected loss is the product of the estimated coefficients and the loss on the market portfolio for that day. The projections under the two approaches are $\hat{L}_{EVT,j} =$

$L_m\hat{\beta}^T_{EVT,j}$ and $\hat{L}_{OLS,j} = L_m\hat{\beta}^T_{OLS,j}$, respectively.[7] We compare the performance of the two approaches by their root mean squared error (RMSE) calculated as $\sqrt{\frac{1}{N}\sum_j (L_j - \hat{L}_{EVT,j})^2}$ and $\sqrt{\frac{1}{N}\sum_j (L_j - \hat{L}_{OLS,j})^2}$. The best-performing method should report a lower RMSE.

Table 2 reports the RMSE of the projected portfolio losses for both approaches in each subperiod. In all subperiods, the EVT approach reports a lower RMSE than the conditional regression approach.[8] The average reduction in the RMSE is approximately 40%. We report t-statistics and p-values in the last two columns to test against the null hypothesis of a larger RMSE from the EVT approach. In 11 out of 16 subperiods, the null is rejected at the 5% significance level.[9] We zoom in on two subperiods with relatively big and relatively small reductions in the RMSE.

A substantial improvement in the projections in terms of RMSE is during the stock market plunge of 8.1% on 18 October 1937. This stock market crash occurred during a period of high uncertainty about the US economy: During the nine-month economic decline from September 1937 to June 1938, national income fell by 12% and firm profits fell by 78%; see e.g., Roose (1948). The errors in the projected losses with the EVT approach are significantly smaller at the 1% significance level.

The improvement in the projections is limited for the stock market plunge of 5.0% on 17 September 2001. This was the day that the NYSE resumed trading after the terrorist attacks six days earlier. For this stock market crash, the RMSE of the regression approach is 5.76, while the RMSE for the EVT approach is 5.18 (no significant difference). A relatively high proportion of the forecasting errors on this day are due to the reaction of a few industries because of the nature of this particular event. The industry portfolios "Defense" and "Shipbuilding" would usually move in the same direction as the stock market index. However, during this crash, they reported a gain of 15% and 7%, respectively. In contrast, because of the nature of the terrorist attacks, the industry portfolios "Aircraft" and "Transportation" (which includes "Air transportation") reacted much more strongly. These portfolios lost 18% and 14%, respectively. The linear tail model does not anticipate the profits and losses

for these four portfolios in this case. If the projection errors for these four portfolios are excluded, the RMSE of the EVT approach and conditional approach decline substantially to a level of 4.40 and 3.48, respectively. Moreover, the difference between the MSE of the EVT approach and that of the conditional regression approach in this subperiod becomes (weakly) significant at the 10% level if these four portfolios are excluded.

To summarize, the EVT approach shows a better overall performance than the conditional regression approach in projecting portfolio losses on the worst market day. We interpret this better performance as resulting from improved accuracy in estimating β^T based on a small number of extreme observations.

4 CONCLUDING REMARKS

In this paper, we propose an EVT approach for estimating β^T in the linear tail model based on only a small number of extreme observations. Simulations show that our EVT approach yields a lower mean squared error than conditional regressions on tail observations. We demonstrate one application of the EVT approach: projecting large losses on stock portfolios during extremely adverse market conditions. Beyond the scope of the current paper, systematic risk under extremely adverse market conditions may have asset pricing implications. An application in that direction can be found in Van Oordt and Zhou (forthcoming). The method can be applied to other variables known to be heavy-tailed, such as trading volume, insurance claims and the severity of natural disasters.

The estimator of β^T might be further improved by considering more sophisticated EVT techniques. For instance, the estimator on the tail index, $\hat{\alpha}_x$, suffers from an asymptotic bias issue. Many studies propose different bias corrected estimators; see, e.g., Peng (1998), Feuerverger and Hall (1999) and Gomes et al. (2008). In addition, Fougères et al. (2015) propose a bias corrected estimator for the τ-measure. Whether the use of such sophisticated

techniques will improve the current simple estimation procedure for β^T offers an interesting perspective for future research.

Notes

[1]The τ-measure in Eq. (2.2) is closely related to the measure $E(\kappa|\kappa \geq 1)$ introduced by Huang (1992) and applied by Hartmann et al. (2004). There, κ is the number of events occurring with probability p and $E(\kappa|\kappa \geq 1)$ is the expected number of tail events given that there is at least one. In the bivariate case, the two measures are connected by $E(\kappa|\kappa \geq 1) = \frac{2}{2-\tau}$.

[2]Formally, the estimator in the conditional regression approach is

$$\hat{\beta}^T_{OLS} = \frac{\sum_{\{t:X_t<X_{n,k+1}\}} \left(Y_t - \bar{Y}^T\right)\left(X_t - \bar{X}^T\right)}{\sum_{\{t:X_t<X_{n,k+1}\}} \left(X_t - \bar{X}^T\right)^2},$$

where $\bar{X}^T = (1/k)\sum_{\{t:X_t<X_{n,k+1}\}}(X_t)$ and $\bar{Y}^T = (1/k)\sum_{\{t:X_t<X_{n,k+1}\}}(Y_t)$. The estimator in the conditional regression approach is theoretically unbiased. A direct theoretical comparison of the asymptotic variances of the EVT approach and the conditional regression approach is difficult because their levels depend on different statistical parameters. For example, the asymptotic variance of $\hat{\beta}^T_{OLS}$ depends on the variance of ε, while the EVT approach does not assume finite variances.

[3]In a sample of 1,250 observations, about $1,250 \times 3.0\% = 37.5$ observations are expected to be generated from the linear tail model.

[4]To illustrate the decomposition of the MSE, we report the squared bias in the figures, which do not show the sign of the bias. Numerical results show that the bias of the EVT approach is consistently positive for all simulated models.

[5]Data and documentation are available from the personal website of Kenneth French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Our results are based on data accessed on 8 January 2014. The secondary data are based on the returns of stocks listed on NYSE, AMEX and NASDAQ in the CRSP database. The definition of the industry portfolios is based on SIC-codes. Industry portfolios with missing returns in a subperiod are excluded from the analysis for that specific subperiod. Five industry portfolios report missing returns in the start of the sample. From July 1963 onwards, one industry portfolio remains unavailable ("Healthcare," SIC-codes 8000-8099). After July 1969, all 48 portfolios are available.

[6]Our results remain qualitatively unchanged when these portfolios are included in the analysis.

[7]Our conclusions remain the same if the projected losses under the conditional regression approach are calculated as $\hat{L}_{OLS,j} = C^T_{OLS,j} + L_m \hat{\beta}^T_{OLS,j}$, where $C^T_{OLS,j} = \bar{Y}^T - \hat{\beta}_{OLS,j}\bar{X}^T$. In this case, the EVT approach provides a significantly lower RMSE at the 5% significance level in 6 out of 16 cases (8 out of 16 cases at the 10% significance level).

[8]In all subperiods, the mean absolute error from the EVT approach is also below that of the conditional regression approach.

[9]The better performance of the EVT approach remains for various levels of k. More specifically, with k fixed at 20, 30, 35, 40, 45 and 50, the null is rejected at the 5% significance level for 11, 11, 9, 8, 5 and 5 subperiods, respectively. In line with our simulation results in Subsection 2.3, the better performance of the EVT approach relative to the conditional regression approach becomes weaker for higher levels of k.

APPENDIX: PROOFS

We start with proving the following lemma, which helps to handle the probability of joint extreme events.

Lemma 1 *Suppose $\beta^T > 0$ and $\alpha_y > \frac{1}{2}\alpha_x$. Under the linear tail model in Eq. (1.1) and the heavy-tailedness of the downside distributions in (2.1), we have that*

$$\lim_{p\to 0} \frac{\Pr\left(Y < Q_y(py), X < Q_x(px)\right)}{\Pr\left(X < \min\left(\frac{Q_y(py)}{\beta^T}, Q_x(px)\right)\right)} = 1, \tag{A.1}$$

uniformly for $(x, y) \in (0, 3/2]^2$.

Proof of Lemma 1. Denote the following sets

$$\begin{aligned}
C &:= \{Y < Q_y(py), X < Q_x(px)\} = \left\{\beta^T X + \varepsilon < Q_y(py), X < Q_x(px)\right\}, \\
C_0 &:= \left\{X < \min\left(\frac{Q_y(py)}{\beta^T}, Q_x(px)\right)\right\}, \\
C_1 &:= \left\{\beta^T X < Q_y(py)(1+\delta), X < Q_x(px), \varepsilon < -\delta Q_y(py)\right\}, \\
C_{21} &:= \left\{\beta^T X < Q_y(py)(1-\delta), X < Q_x(px)\right\}, \\
C_{22} &:= \left\{\varepsilon < \delta Q_y(py), X < Q_x(px)\right\},
\end{aligned}$$

where $\delta := \delta(p) = p^c$, with $0 < c < \min\left(\frac{1}{\alpha_y}, \frac{1}{\alpha_y}\left(2 - \frac{\alpha_x}{\alpha_y}\right)\right)$. Here we use the assumption that $\alpha_y > \frac{1}{2}\alpha_x$. Since the distribution function of Y follows (2.1), we get that $Q_y(p)$ is a regularly varying function with index $-1/\alpha_y$ as $p \to 0$, which leads to the regular variation of $\delta(p)Q_y(p)$ with index $c - 1/\alpha_y < 0$. Consequently, $\delta(p) \to 0$ and $\delta(p)Q_y(p) \to -\infty$ as $p \to 0$. This further implies that $\delta(p)Q_y(py) \to -\infty$ uniformly for $0 < y \leq 3/2$ as $p \to \infty$.

It is obvious that $C_1 \subset C \subset C_{21} \bigcup C_{22}$. Therefore, as $p \to 0$, to prove that $\frac{\Pr(C)}{\Pr(C_0)} \to 1$, we show that

$$\frac{\Pr(C_1)}{\Pr(C_0)} \to 1, \quad \frac{\Pr(C_{21})}{\Pr(C_0)} \to 1 \quad \text{and} \quad \frac{\Pr(C_{22})}{\Pr(C_0)} \to 0.$$

We first deal with C_1. Since X and ε are independent, we get that

$$\Pr(C_1) = \Pr\left(X < \min\left(\frac{Q_y(py)}{\beta^T}(1+\delta(p)), Q_x(px)\right)\right) \Pr\left(\varepsilon < -\delta(p)Q_y(py)\right).$$

Since $-\delta(p)Q_y(py) \to +\infty$ as $p \to 0$, the second term $\Pr\left(\varepsilon < -\delta(p)Q_y(py)\right)$ tends to 1 as $p \to 0$. Therefore, we need to show that

$$\lim_{p\to 0} \frac{\Pr\left(X < \min\left(\frac{Q_y(py)}{\beta^T}(1+\delta(p)), Q_x(px)\right)\right)}{\Pr\left(X < \min\left(\frac{Q_y(py)}{\beta^T}, Q_x(px)\right)\right)} = 1. \tag{A.2}$$

This relation follows directly from the combination of the following three facts: first, we have inequalities

$$(1+\delta(p)) \min\left(\frac{Q_y(py)}{\beta^T}, Q_x(px)\right) \le \min\left(\frac{Q_y(py)}{\beta^T}(1+\delta(p)), Q_x(px)\right) \le \min\left(\frac{Q_y(py)}{\beta^T}, Q_x(px)\right);$$

second, the distribution function of X is regularly varying at $-\infty$; third, $\delta(p) \to 0$ as $p \to 0$. It is obvious that the convergence in (A.2) holds uniformly for all $(x,y) \in (0,3/2]^2$. In a similar way, one can show that $\frac{\Pr(C_{21})}{\Pr(C_0)} \to 1$ uniformly for all $(x,y) \in (0,3/2]^2$, as $p \to 0$.

Last, we deal with C_{22}. With set manipulation and the independence between X and ε, we get that

$$\begin{aligned}
&\Pr(\varepsilon < \delta(p)Q_y(py)) \Pr\left(X < -\frac{\delta(p)}{2}Q_y(py)\right) \\
=&\Pr\left(\varepsilon < \delta(p)Q_y(py), X < -\frac{\delta(p)}{2}Q_y(py)\right) \\
\le&\Pr(Y < \frac{\delta(p)}{2}Q_y(py)).
\end{aligned}$$

Using the properties of regularly varying functions, we derive that $\Pr(Y < \frac{\delta(p)}{2}Q_y(py))$ is a regularly varying function as $p \to 0$ with index $(c - 1/\alpha_y)(-\alpha_y) = 1 - c\alpha_y > 0$. Thus, for any given $\eta > 0$, with sufficiently small p, we get that $\Pr(Y < \frac{\delta(p)}{2}Q_y(py)) \le p^{1-c\alpha_y-\eta}$. It

implies that

$$\limsup_{p\to\infty}\frac{\Pr(C_{22})}{p^{2-c\alpha_y-\eta}}=\limsup_{p\to\infty}\frac{\Pr(\varepsilon<\delta(p)Q_y(py))}{p^{1-c\alpha_y-\eta}}x\le 2\limsup_{p\to\infty}\frac{\Pr(Y<\frac{\delta(p)}{2}Q_y(py))}{p^{1-c\alpha_y-\eta}}\le 2$$

holds uniformly for all $(x,y)\in(0,3/2]^2$.

On the other hand,

$$\Pr(C_0)=\min\left(\Pr\left(X<\min\left(\frac{Q_y(py)}{\beta^T}\right)\right),px\right).$$

Again, by using the properties of regularly varying functions, we derive that $\Pr\left(X<\min\left(\frac{Q_y(py)}{\beta^T}\right)\right)$ is a regularly varying function as $p\to 0$ with index α_x/α_y. Notice that c satisfies the inequality $c<\min\left(\frac{1}{\alpha_y},\frac{1}{\alpha_y}\left(2-\frac{\alpha_x}{\alpha_y}\right)\right)$, i.e., $2-c\alpha_y>\max\left(\frac{\alpha_x}{\alpha_y},1\right)$. Thus, we can choose η such that $2-c\alpha_y-\eta>\max\left(\frac{\alpha_x}{\alpha_y},1\right)$. With such a choice, we get that

$$\lim_{p\to\infty}\frac{\Pr(C_0)}{p^{2-c\alpha_y-\eta}}=+\infty.$$

Therefore, we can compare $\Pr(C_{22})$ and $\Pr(C_0)$ to get that

$$\lim_{p\to 0}\frac{\Pr(C_{22})}{\Pr(C_0)}=0,$$

which completes the proof of the lemma. ■

Proof of Theorem 1. If $\beta^T=0$, then $\tau(p)=p$. Since the distribution function of X and Y are regularly varying, we get that $Q_x(p)$ and $Q_y(p)$ are regularly varying as $p\to 0$, with indices $-1/\alpha_x$ and $-1/\alpha_y$ respectively. Consequently, $(\tau(p))^{1/\alpha_x}\frac{Q_y(p)}{Q_x(p)}$ is a regularly varying function with index $2/\alpha_x-1/\alpha_y$ as $p\to 0$. The condition $\alpha_y>\alpha_x/2$ implies that $2\alpha_x-1/\alpha_y>0$, which leads to

$$\lim_{p\to 0}(\tau(p))^{1/\alpha_x}\frac{Q_y(p)}{Q_x(p)}=0=\beta^T.$$

If $\beta^T > 0$, we first show that

$$Q_y(p) \leq \beta^T Q_x(p) \tag{A.3}$$

for sufficiently small p.

Due to the heavy-tailedness, we have that $Q_x(p)$ converges to minus infinity as $p \to 0$. Thus, when the tail probability p is sufficiently low, such that $Q_x(p)$ is smaller than the threshold in the linear tail model, the linear relation in Eq. (1.1) is valid for $X < Q_x(p)$. Hence we have that for any $\delta > 0$,

$$\begin{aligned} \Pr(Y < \beta^T Q_x(p)) &\geq \Pr(\beta^T X < \beta^T Q_x(p)(1+\delta), \varepsilon < -\delta\beta^T Q_x(p)) \\ &= \Pr(X < Q_x(p)(1+\delta)) \Pr(\varepsilon < -\delta\beta^T Q_x(p)). \end{aligned} \tag{A.4}$$

The last step is due to the independence between X and ε. Since the distribution function of X is regularly varying at $-\infty$ with index $-\alpha_x$, we get that

$$\lim_{p\to 0} \frac{\Pr(X < Q_x(p)(1+\delta))}{\Pr(X < Q_x(p))} = (1+\delta)^{-\alpha_x}.$$

Moreover, $\lim_{p\to 0} \Pr(\varepsilon < -\delta\beta^T Q_x(p)) = 1$. Thus, from (A.4), we obtain

$$\liminf_{p\to 0} \frac{\Pr(Y < \beta^T Q_x(p))}{p} \geq (1+\delta)^{-\alpha_x}.$$

Notice that the above inequality holds for any $\delta > 0$. By taking $\delta \to 0$, we obtain that

$$\liminf_{p\to 0} \frac{\Pr(Y < \beta^T Q_x(p))}{p} \geq 1.$$

Since the quantile $Q_y(p)$ is defined as $\Pr(Y < Q_y(p)) = p$, we obtain that the inequality in (A.3) holds for sufficiently low probability p.

Next, we apply Lemma 1 with $x = y = 1$ and the fact that $\min\left(\frac{Q_y(p)}{\beta^T}, Q_x(p)\right) = \frac{Q_y(p)}{\beta^T}$. We get

$$\lim_{p\to 0} \frac{p\tau(p)}{\Pr\left(X < \frac{Q_y(p)}{\beta^T}\right)} = 1.$$

Following the regularly varying tail of X, we get that

$$\lim_{p\to 0} \frac{\Pr\left(X < \frac{Q_y(p)}{\beta^T}\right)}{\Pr(X < Q_x(p))} \cdot \left(\frac{Q_y(p)}{\beta^T Q_x(p)}\right)^{\alpha_x} = 1.$$

By combining the two limit relations above, we get that

$$\lim_{p\to 0} \tau(p) \cdot \left(\frac{Q_y(p)}{\beta^T Q_x(p)}\right)^{\alpha_x} = 1,$$

which completes the proof of the theorem. ■

Proof of Theorem 2. Write

$$\hat{\beta}^T = \left(\frac{\hat{\tau}(k/n)}{\tau(k/n)}\right)^{1/\hat{\alpha}_x} \cdot (\tau(k/n))^{1/\hat{\alpha}_x - 1/\alpha_x} \cdot \frac{\hat{Q}_y(k/n)}{Q_y(k/n)} \cdot \frac{Q_x(k/n)}{\hat{Q}_x(k/n)} \cdot \left((\tau(k/n))^{1/\alpha_x} \frac{Q_y(k/n)}{Q_x(k/n)}\right)$$
$$=: I_1 \cdot I_2 \cdot I_3 \cdot I_4 \cdot I_5.$$

The classic consistency results in extreme value statistics ensures that $\hat{\alpha}_x \xrightarrow{P} \alpha_x$, $\frac{\hat{Q}_x(k/n)}{Q_x(k/n)} \xrightarrow{P} 1$ and $\frac{\hat{Q}_y(k/n)}{Q_y(k/n)} \xrightarrow{P} 1$ as $n \to \infty$; see Theorem 3.2.2 and Corollary 4.3.9 in De Haan and Ferreira (2006). Hence $I_3, I_4 \xrightarrow{P} 1$ as $n \to \infty$. Theorem 1 ensures that $I_5 \to \beta^T$ as $n \to \infty$. Therefore, the only issues that are left to prove are $I_1, I_2 \xrightarrow{P} 1$ as $n \to \infty$.

We first deal with I_1, which is equivalent to prove the consistency of $\hat{\tau}(k/n)$. Denote

$$\tilde{\tau}(x,y) = \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}_{\left\{X_i < Q_x\left(\frac{k}{n}x\right) \text{ and } Y_i < Q_y\left(\frac{k}{n}y\right)\right\}}, \tag{A.5}$$

for (x,y) in the neighborhood of $(1,1)$. Then $\hat{\tau}(k/n)$ can be written as

$$\hat{\tau}(k/n) = \tilde{\tau}\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}), \frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1})\right).$$

Here $\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}), \frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1})\right)$ is in the neighborhood of $(1,1)$ in the following sense. According to Corollary 2.2.2 in De Haan and Ferreira (2006), as $n \to \infty$,

$$\sqrt{k}\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}) - 1\right) \xrightarrow{d} N(0,1).$$

Hence for any $\delta > 0$, as $n \to \infty$,

$$\Pr\left(\left|\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}) - 1\right| > k^{-1/2+\delta}\right) \to 0.$$

A similar relation for $Y_{n,k+1}$ holds. Therefore, it suffices to consider $\tilde{\tau}(x,y)$ for $(x,y) \in [1-k^{-1/2+\delta}, 1+k^{-1/2+\delta}]^2$ for some $0 < \delta < 1/2$.

By applying the law of large number, we get that as $n \to \infty$,

$$\frac{\tilde{\tau}(x,y)}{R(x,y,k/n)} \xrightarrow{P} 1, \tag{A.6}$$

where $R(x,y,k/n) := \frac{n}{k}\Pr\left(X < Q_x\left(\frac{k}{n}x\right), Y < Q_y\left(\frac{k}{n}y\right)\right)$. Notice that $\tau(k/n) = R(1,1,k/n)$.

In order to replace the denominator in (A.6) by $\tau(k/n)$, we need to prove that

$$\lim_{n\to\infty}\frac{R(x,y,k/n)}{\tau(k/n)} = 1 \tag{A.7}$$

holds uniformly for all $(x,y) \in [1-k^{-1/2+\delta}, 1+k^{-1/2+\delta}]^2$.

If $\beta^T = 0$, as $n \to \infty$, for all $(x,y) \in [1-k^{-1/2+\delta}, 1+k^{-1/2+\delta}]^2$, we have uniformly

$$\lim_{n\to\infty}\frac{R(x,y,k/n)}{\tau(k/n)} = \lim_{n\to\infty}\frac{\frac{n}{k}\frac{k}{n}x\frac{k}{n}y}{\left(\frac{k}{n}\right)^2} = \lim_{n\to\infty} xy = 1.$$

If $\beta^T > 0$, we apply Lemma 1 with $p = k/n$ and get that the relation

$$\lim_{n\to\infty} \frac{R(x,y,k/n)}{\frac{n}{k}\Pr\left(X < \min\left(\frac{Q_y\left(\frac{k}{n}y\right)}{\beta^T}, Q_x\left(\frac{k}{n}x\right)\right)\right)} = 1$$

holds uniformly for $(x,y) \in [1-k^{-1/2+\delta}, 1+k^{-1/2+\delta}]^2 \subset (0,3/2]^2$.

We apply the regularly varying properties of the distribution and quantile function of X and Y to further simplify the denominator. We get that, as $n \to \infty$,

$$\frac{n}{k}\Pr\left(X < \min\left(\frac{Q_y\left(\frac{k}{n}y\right)}{\beta^T}, Q_x\left(\frac{k}{n}x\right)\right)\right) = \frac{\Pr\left(X < \min\left(\frac{Q_y\left(\frac{k}{n}y\right)}{\beta^T}, Q_x\left(\frac{k}{n}x\right)\right)\right)}{\Pr\left(X < Q_x\left(\frac{n}{k}\right)\right)}$$

$$\sim \left(\frac{\min\left(\frac{Q_y\left(\frac{k}{n}y\right)}{\beta^T}, Q_x\left(\frac{k}{n}x\right)\right)}{Q_x\left(\frac{n}{k}\right)}\right)^{-\alpha_x} = \min\left(\left(\frac{\beta^T Q_x\left(\frac{n}{k}\right)}{Q_y\left(\frac{k}{n}y\right)}\right)^{\alpha_x}, \left(\frac{Q_x\left(\frac{n}{k}\right)}{Q_x\left(\frac{k}{n}x\right)}\right)^{\alpha_x}\right).$$

From Theorem 1, we get that $\left(\frac{\beta^T Q_x\left(\frac{n}{k}\right)}{Q_y\left(\frac{k}{n}y\right)}\right)^{\alpha_x} \sim \tau(k/n)$ as $n \to \infty$ holds uniformly for $|y-1| \le k^{-1/2+\delta}$. In addition, $\left(\frac{Q_x\left(\frac{n}{k}\right)}{Q_x\left(\frac{k}{n}x\right)}\right)^{\alpha_x} \to 1$ as $n \to \infty$ holds uniformly for $|x-1| \le k^{-1/2+\delta}$. Together with $\tau(k/n) \le 1$, we get that

$$\frac{n}{k}\Pr\left(X < \min\left(\frac{Q_y\left(\frac{k}{n}y\right)}{\beta^T}, Q_x\left(\frac{k}{n}x\right)\right)\right) \sim \tau(k/n)$$

holds uniformly for $(x,y) \in [1-k^{-1/2+\delta}, 1+k^{-1/2+\delta}]^2$, as $n \to \infty$. Hence, we proved the equation (A.7).

By combining (A.6) and (A.7), we get that for all $(x,y) \in [1-k^{-1/2+\delta}, 1+k^{-1/2+\delta}]^2$, we can replace the denominator in (A.6) by $\tau(k/n)$. After that, we can apply it to the random location $\left(\frac{n}{k}Q_x^{\leftarrow}\left(X_{n,k+1}\right), \frac{n}{k}Q_y^{\leftarrow}\left(Y_{n,k+1}\right)\right) \in [1-k^{-1/2+\delta}, 1+k^{-1/2+\delta}]^2$ and obtain the

consistency for the $\hat{\tau}(k/n)$ estimator: as $n \to \infty$,

$$\frac{\hat{\tau}(k/n)}{\tau(k/n)} \xrightarrow{P} 1,$$

which guarantees that $I_1 \xrightarrow{P} 1$. We remark that our proof allows for $\lim_{p\to 0}\tau(p) = 0$, which goes beyond the typical consistency results in bivariate extreme value statistics.

Finally, we deal with I_2. If $\limsup_{p\to 0}\tau(p) > 0$, then the consistency of $\hat{\alpha}_x$ leads to $I_2 \xrightarrow{P} 1$, as $n \to \infty$. In this case, the theorem is proved without using conditions (2.7)–(2.9).

If $\lim_{p\to 0}\tau(p) = 0$, to prove $I_2 \xrightarrow{P} 1$, we need to prove that as $n \to \infty$,

$$\log\tau(k/n)\left(\frac{1}{\hat{\alpha}_x} - \frac{1}{\alpha_x}\right) \xrightarrow{P} 0. \tag{A.8}$$

The conditions (2.7) and (2.8) imply the asymptotic normality for $\hat{\alpha}_x$: as $n \to \infty$, $\sqrt{k_1}\left(\frac{1}{\hat{\alpha}_x} - \frac{1}{\alpha_x}\right) = O_p(1)$; see, e.g., Theorem 3.2.5 in De Haan and Ferreira (2006). Therefore, it only remains to prove that $\log\tau(k/n) = o(\sqrt{k_1})$ as $n \to \infty$.

If $\beta^T = 0$, then $\tau(k/n) = k/n > 1/n$. Hence, as $n \to \infty$, $\log\tau(k/n) = O(\log n)$. If $\beta^T > 0$, following Theorem 1, we get that for sufficiently large n,

$$\tau(k/n) \sim \left(\beta^T \frac{Q_x\left(\frac{k}{n}\right)}{Q_y\left(\frac{k}{n}\right)}\right)^{\alpha_x} > D\left(\frac{k}{n}\right)^{\alpha_x/\alpha_y - 1 + \delta},$$

for some $D > 0$ and $\delta > 0$. Here, the last step comes from the Potter inequality for regularly varying function; see, e.g., inequality (B.1.19) in De Haan and Ferreira (2006). Again, we get that $\log\tau(k/n) = O(\log n)$ as $n \to \infty$.

Combining $\log\tau(k/n) = O(\log n)$ as $n \to \infty$ with the condition (2.9), we have that $\log\tau(k/n) = o(\sqrt{k_1})$ as $n \to \infty$, which implies that $I_2 \xrightarrow{P} 1$ holds also for the case $\lim_{p\to 0}\tau(p) = 0$. ■

Proof of Theorem 3. We start by deriving the explicit form for $R(x, y)$ and its partial derivatives at $(1, 1)$ because these quantities play an important role in the calculation of the

asymptotic variance. Notice that R is a homogeneous function with degree 1. Thus, it is only necessary to derive $R(x,1)$ for $x > 0$. This is given in the following lemma.

Lemma 2 *Under the conditions in Theorem 3, we have* $R(x,1) = \min(x,\tau)$ *for* $x > 0$.

Proof. If $\tau = 1$, we get that X and Y are completely tail dependent. Consequently, $R(x,1) = \min(x,1)$. The lemma is proved for this case.

Next, we handle the case $\tau < 1$. Theorem 1 implies that

$$\lim_{p\to 0} \frac{Q_y(p)}{Q_x(\tau p)} = \lim_{p\to 0} \tau^{1/\alpha} \frac{Q_y(p)}{Q_x(p)} = \beta^T.$$

Hence, for any $\tau < x < 1$, we have that for sufficiently small p, $Q_y(p) < \beta^T Q_x(xp)$. On the other hand, for any $0 < x < \tau$, for sufficiently small p, $Q_y(p) > \beta^T Q_x(xp)$.

By applying Lemma 1 with $y = 1$, we get that

$$\lim_{p\to 0} \frac{R(x,1)}{\frac{1}{p}\Pr\left(X < \min\left(\frac{Q_y(p)}{\beta^T}, Q_x(px)\right)\right)} = 1.$$

In particular, for $\tau < x < 1$, since $Q_y(p) < \beta^T Q_x(xp)$,

$$R(x,1) = \lim_{p\to 0} \frac{\Pr(\beta^T X < Q_y(p))}{\Pr(X < Q_x(p))} = \lim_{p\to 0} \left(\frac{Q_y(p)}{\beta^T Q_x(p)}\right)^{-\alpha_x} = \tau.$$

On the other hand, for $0 < x < \tau$,

$$R(x,1) = \lim_{p\to 0} \frac{\Pr(X < Q_x(xp)}{p} = x.$$

Finally, for the point $x = \tau$, we use continuity of the $R(x,1)$ function to get that $R(\tau,1) = \tau$. The lemma is thus proved. ■

As a direct consequence of Lemma 2, we get that for $\frac{x}{y} > \tau$

$$R(x,y) = yR(\frac{x}{y},1) = \tau y.$$

Hence $R_1(1,1) = 0$ and $R_2(1,1) = \tau$, where R_1, R_2 denotes the partial derivatives of R with respect to x and y respectively.

By interchanging partial derivatives with taking the limit, we get that, as $n \to \infty$,

$$\frac{\partial}{\partial x} R(x,y,k/n)|_{\{(1,1)\}} \to R_1(1,1) = 0 \text{ and } \frac{\partial}{\partial y} R(x,y,k/n)|_{\{(1,1)\}} \to R_2(1,1) = \tau.$$

Now we can deal with the asymptotic normality of the estimator $\hat{\beta}^T$ by employing the asymptotic normality results for the four elements in the literature. In particular, we target to get the covariance matrix of their asymptotic limit. This is achieved by studying the asymptotic behavior of the process $\bar{\tau}(x,y)$ defined in (A.5).

With the second order condition (2.11) and the fact that $k = o(n^{2\theta/(1+2\theta)})$, we can apply Proposition 3.1 in Einmahl et al. (2006) and obtain that

$$\sup_{(x,y)\in[0,T]^2/\{(0,0)\}} \frac{1}{(\max(x,y))^\lambda} \left| \sqrt{k} \left(\bar{\tau}(x,y) - R(x,y,k/n) \right) - W(x,y) \right| \xrightarrow{P} 0, \tag{A.9}$$

for given $T > 0$ and $0 \leq \lambda < 1/2$, where $W(x,y)$ is a continuous mean zero Gaussian process with the following covariance structure:

$$\mathrm{E}\, W(x_1,y_1)W(x_2,y_2) = R(\min(x_1,x_2), \min(y_1,y_2)).$$

In addition, for marginals, we have that

$$\sup_{0<x\leq T} \frac{1}{x^\lambda} \left| \sqrt{k} \left(\frac{1}{k} \sum_{i=1}^{n} \mathbf{1}_{\left\{ X_i < Q_x\left(\frac{k}{n}x\right) \right\}} - x \right) - W(x,+\infty) \right| \xrightarrow{P} 0, \tag{A.10}$$

and

$$\sup_{0<y\leq T} \frac{1}{y^\lambda} \left| \sqrt{k} \left(\frac{1}{k} \sum_{i=1}^{n} \mathbf{1}_{\left\{ Y_i < Q_y\left(\frac{k}{n}y\right) \right\}} - y \right) - W(+\infty,y) \right| \xrightarrow{P} 0. \tag{A.11}$$

By taking $(x,y) = \left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}), \frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1})\right)$ in (A.9), we get that, as $n \to \infty$,

$$\sqrt{k}\left(\hat{\tau}(k/n) - R\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}), \frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1}), \frac{k}{n}\right)\right) - W\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}), \frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1})\right) \xrightarrow{P} 0.$$

Following the continuity of the sample path for the W process, we get that, as $n \to \infty$,

$$W\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}), \frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1})\right) \xrightarrow{P} W(1,1).$$

Hence, we get that

$$\sqrt{k}\left(\hat{\tau}(k/n) - R\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}), \frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1}), \frac{k}{n}\right)\right) \xrightarrow{P} W(1,1). \tag{A.12}$$

By writing

$$\begin{aligned}
&R\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}), \frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1}), \frac{k}{n}\right) \\
=&R(1,1,k/n) \\
&+ \frac{\partial}{\partial x}R(x,y,k/n)|_{\{(1,1)\}}\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}) - 1\right) + \frac{\partial}{\partial y}R(x,y,k/n)|_{\{(1,1)\}}\left(\frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1}) - 1\right) \\
&+ o\left(\left|\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}) - 1\right| + \left|\frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1}) - 1\right|\right),
\end{aligned}$$

we get that as, $n \to \infty$,

$$\begin{aligned}
&\sqrt{k}\left(R\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}), \frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1}), \frac{k}{n}\right) - R(1,1,k/n)\right) \\
=&\frac{\partial}{\partial x}R(x,y,k/n)|_{\{(1,1)\}}\sqrt{k}\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}) - 1\right) + \frac{\partial}{\partial y}R(x,y,k/n)|_{\{(1,1)\}}\sqrt{k}\left(\frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1}) - 1\right) \\
&+ o\left(\sqrt{k}\left|\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1}) - 1\right| + \sqrt{k}\left|\frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1}) - 1\right|\right) \\
\xrightarrow{P}& -\tau W(+\infty, 1).
\end{aligned} \tag{A.13}$$

Here in the last step, we used the limits of the two partial derivatives as well as the facts that

$$\sqrt{k}\left(\frac{n}{k}Q_x^{\leftarrow}(X_{n,k+1})-1\right)\xrightarrow{P}-W(1,+\infty) \text{ and } \sqrt{k}\left(\frac{n}{k}Q_y^{\leftarrow}(Y_{n,k+1})-1\right)\xrightarrow{P}-W(\infty,+1),$$

as $n\to\infty$. The latter two asymptotic relations can be derived from inverting the marginal asymptotic properties in (A.10) and (A.11) using the Vervaat lemma; see Vervaat (1972).

By combining (A.12) and (A.13), we get that, as $n\to\infty$,

$$\sqrt{k}\left(\frac{\hat{\tau}(k/n)}{\tau(k/n)}-1\right)\xrightarrow{P}\frac{1}{\tau}W(1,1)-W(+\infty,1). \tag{A.14}$$

Next, by applying the asymptotic properties for marginals in (A.10) and (A.11), we get that, as $n\to\infty$,

$$\sqrt{k}\left(\frac{1}{\hat{\alpha}_x}-\frac{1}{\alpha_x}\right)\xrightarrow{P}\frac{1}{\alpha_x}\left(\int_0^1 W(s,+\infty)\frac{ds}{s}-W(1,+\infty)\right), \tag{A.15}$$

$$\sqrt{k}\left(\frac{\hat{Q}_x(k/n)}{Q_x(k/n)}-1\right)\xrightarrow{P}\frac{1}{\alpha_x}W(1,+\infty), \tag{A.16}$$

$$\sqrt{k}\left(\frac{\hat{Q}_y(k/n)}{Q_y(k/n)}-1\right)\xrightarrow{P}\frac{1}{\alpha_y}W(+\infty,1). \tag{A.17}$$

Here we use the conditions $k=o(n^{\zeta})$ with $\zeta<\min\left(\frac{2\gamma}{\alpha_x+2\gamma},\frac{2\gamma'}{\alpha_y+2\gamma'}\right)$. For the derivation of the three relations, see Example 5.1.5, and equation (5.1.19) in De Haan and Ferreira (2006).

Using Cramér's delta method, we can assemble the asymptotic relation in (A.14)–(A.17) to obtain that

$$\sqrt{k}\left(\frac{\hat{\beta}^T}{(\tau(k/n))^{1/\alpha_x}\frac{Q_y(k/n)}{Q_x(k/n)}}-1\right)\xrightarrow{P}\Gamma$$

where

$$\begin{aligned}\Gamma &= \frac{1}{\alpha_x}\left(\frac{1}{\tau}W(1,1) - W(+\infty,1)\right) + (\log\tau)\frac{1}{\alpha_x}\left(\int_0^1 W(s,+\infty)\frac{ds}{s} - W(1,+\infty)\right) \\ &\quad + \frac{1}{\alpha_y}W(+\infty,1) - \frac{1}{\alpha_x}W(1,+\infty) \\ &= \frac{1}{\alpha_x}\left(\frac{1}{\tau}W(1,1) + (\log\tau)\int_0^1 W(s,+\infty)\frac{ds}{s} - (1+\log\tau)W(1,+\infty)\right).\end{aligned}$$

Using the expression of $R(x,1)$ in Lemma 2, one can calculate that

$$\mathrm{Var}(\Gamma) = \frac{1}{\alpha_x^2}\left(\frac{1}{\tau} - 1 - (\log\tau)^2\right).$$

Therefore, what remains to be proved is the following deterministic relation

$$\lim_{n\to\infty}\sqrt{k}\left(\frac{(\tau(k/n))^{1/\alpha_x}\frac{Q_y(k/n)}{Q_x(k/n)}}{\beta^T} - 1\right) = 0.$$

Knowing that $\lim_{n\to\infty}\frac{Q_y(k/n)}{Q_x(k/n)} = \beta^T\tau^{-1/\alpha_x} > 0$, the above relation is equivalent to

$$\lim_{n\to\infty}\sqrt{k}\left(\tau(k/n) - \left(\beta^T\frac{Q_x(k/n)}{Q_y(k/n)}\right)^{\alpha_x}\right) = 0. \tag{A.18}$$

To prove (A.18), we revisit the proof of Lemma 1 with considering $p = k/n$, $x = y = 1$. In this case, the denominator is simplified to $\Pr(X < Q_y(k/n)/\beta^T)$. Instead of the $\delta(p)$ considered in the proof, we redefine the δ-term as $\delta_n = k^{-1/2-\delta}$ with $\delta > 0$ such that $(1/2+\delta)\alpha_y + 3/2 < 1/\zeta$. Notice that, since $\zeta < \frac{2}{\alpha_y+3}$, the choice of δ is feasible.

With such a choice, we have that $\lim_{n\to\infty}\sqrt{k}\delta_n = 0$ and $\lim_{n\to\infty}\sqrt{k}\delta_n^{\alpha_y}\frac{k}{n} = 0$. These two relations will be used in the proof of the following limit relations: as $n\to\infty$,

$$\sqrt{k}\Pr(\varepsilon < \delta_n Q_y(k/n)) \to 0 \text{ and } \sqrt{k}\Pr(\varepsilon \geq \delta_n Q_y(k/n)) \to 0, \tag{A.19}$$

$$\sqrt{k}\left(\frac{n}{k}\Pr\left(\beta^T X < Q_y(k/n)(1\pm\delta_n)\right) - \left(\frac{\beta^T Q_x(k/n)}{Q_y(k/n)}\right)^{\alpha_x}\right) = 0. \tag{A.20}$$

Proof of (A.19). We start with the first half. Notice that

$$\begin{aligned}\Pr(\varepsilon < \delta_n Q_y(k/n)) &= \frac{\Pr(\varepsilon < \delta_n Q_y(k/n), X < Q_x(\bar{p}))}{\Pr(X < Q_x(\bar{p}))} \\ &\leq \frac{\Pr(Y < \delta_n Q_y(k/n) + \beta^T Q_x(\bar{p}))}{\bar{p}} \\ &\sim \delta_n^{-\alpha_y} \frac{k/n}{\bar{p}}.\end{aligned}$$

In the last step, we use the regularly varying tail of Y. Since $\sqrt{k}\delta_n^{-\alpha_y}\frac{k}{n} \to 0$ as $n \to \infty$, the first half of (A.19) is proved.

For the second half, we write

$$\begin{aligned}\Pr(\varepsilon \geq -\delta_n Q_y(k/n)) &= \frac{\Pr\left(\varepsilon \geq -\delta_n Q_y(k/n), \frac{1}{\beta^T+1}\delta_n Q_y(k/n) \leq X < Q_x(\bar{p})\right)}{\Pr\left(\frac{1}{\beta^T+1}\delta_n Q_y(k/n) \leq X < Q_x(\bar{p})\right)} \\ &\leq \frac{\Pr\left(Y \geq -\frac{1}{\beta^T+1}\delta_n Q_y(k/n)\right)}{\bar{p} - \Pr\left(X < \frac{1}{\beta^T+1}\delta_n Q_y(k/n)\right)} \\ &\leq D\frac{\Pr\left(Y < \frac{1}{\beta^T+1}\delta_n Q_y(k/n)\right)}{\bar{p} - \Pr\left(X < \frac{1}{\beta^T+1}\delta_n Q_y(k/n)\right)},\end{aligned}$$

for some constant $D > 0$. Here, the last step uses the condition that $\Pr(Y > u) = O(\Pr(Y < -u))$. Notice that the denominator converges to $\bar{p}$, which is positive and finite. The second half of (A.19) is thus proved similar to the proof for the first half. ■

Proof of (A.20). Recall the second order condition (2.7). The condition that $k = O(n^\zeta)$ with $\zeta < \frac{2\gamma}{2\gamma+\alpha_x}$ implies that $\sqrt{k}\eta(Q_x(k/n)) \to 0$. Together with the fact that $\frac{Q_y(k/n)(1+\delta_n)}{\beta^T Q_x(k/n)} \to \tau^{-1/\alpha_x}$, we get that

$$\lim_{n\to\infty} \sqrt{k}\left(\frac{n}{k}\Pr\left(\beta^T X < Q_y(k/n)(1+\delta_n)\right) - \left(\frac{Q_y(k/n)(1+\delta_n)}{\beta^T Q_x(k/n)}\right)^{-\alpha_x}\right) = 0.$$

Therefore (A.20) is proved because $\lim_{n\to\infty}\sqrt{k}\delta_n = 0$. ■

Now we go back to prove (A.18). Recall the sets C, C_0, C_1, C_{21} and C_{22} in the proof of Lemma 1 with $x = y = 1$, $p = k/n$ and $\delta = \delta_n$. Since the relation $C_1 \subset C \subset C_{21} \bigcup C_{22}$ remains valid, we again deal with the three sets that help to bound $\Pr(C)$.

First, the limit relation (A.20) implies that

$$\lim_{n\to\infty} \sqrt{k} \left(\frac{n}{k} \Pr(C_{21}) - \left(\frac{\beta^T Q_x(k/n)}{Q_y(k/n)} \right)^{\alpha_x} \right) = 0.$$

The first half of the limit relation (A.19) implies that $\lim_{n\to\infty} \sqrt{k} \frac{n}{k} \Pr(C_{22}) = 0$. Combining these two, we get that

$$\lim_{n\to\infty} \sqrt{k} \left(\frac{n}{k} \left(\Pr(C_{21}) + \Pr(C_{22}) \right) - \left(\frac{\beta^T Q_x(k/n)}{Q_y(k/n)} \right)^{\alpha_x} \right) = 0.$$

Next, for C_1, due to independency, we have that

$$\Pr(C_1) = \Pr\left(\beta^T X < Q_y(k/n)(1+\delta_n) \right) \cdot \Pr\left(\varepsilon < -\delta_n Q_y(k/n) \right).$$

The second half of the limit relation (A.19) implies that

$$\lim_{n\to\infty} \sqrt{k} \left(\Pr\left(\varepsilon < -\delta_n Q_y(k/n) \right) - 1 \right) = 0.$$

Together with (A.20), we get that

$$\lim_{n\to\infty} \sqrt{k} \left(\frac{n}{k} \Pr(C_1) - \left(\frac{\beta^T Q_x(k/n)}{Q_y(k/n)} \right)^{\alpha_x} \right) = 0.$$

By combining the lower and upper bounds of $\Pr(C)$, we proved (A.18) and thus the theorem.

■

References

A. Ang and J. Chen. Asymmetric Correlations of Equity Portfolios. *Journal of Financial Economics*, 63(3):443–494, 2002.

T. Bollerslev, V. Todorov, and S.Z. Li. Jump Tails, Extreme Dependencies, and the Distribution of Stock Returns. *Journal of Econometrics*, 172(2):307–324, 2013.

L. De Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, New York, 2006.

L. De Haan and U. Stadtmüller. Generalized Regular Variation of Second Order. *Journal of the Australian Mathematical Society (Series A)*, 61(03):381–395, 1996.

C.G. De Vries. The Simple Economics of Bank Fragility. *Journal of Banking & Finance*, 29 (4):803–825, 2005.

H. Drees. Some Aspects of Extreme Value Statistics Under Serial Dependence. *Extremes*, 11(1):35–53, 2008.

J.H.J. Einmahl, L. de Haan, and D. Li. Weighted Approximations of Tail Copula Processes with Application to Testing the Bivariate Extreme Value Condition. *Annals of Statistics*, 34(4):1987–2014, 2006.

P. Embrechts, L. De Haan, and X. Huang. Modelling Multivariate Extremes. In P. Embrechts, editor, *Extremes and Integrated Risk Management*, pages 59–67. RISK Books, London, 2000.

A. Feuerverger and P. Hall. Estimating a Tail Exponent by Modelling Departure From a Pareto Distribution. *Annals of Statistics*, 27(2):760–781, 1999.

A.-L. Fougères, L. de Haan, and C. Mercadier. Bias Correction in Multivariate Extremes. *Annals of Statistics*, 43(2):903–934, 2015.

M.I. Gomes, L. de Haan, and L.H. Rodrigues. Tail Index Estimation for Heavy-Tailed Models: Accommodation of Bias in Weighted Log-Excesses. *Journal of the Royal Statistical Society: Series B*, 70(1):31–52, 2008.

P. Hartmann, S. Straetmans, and C.G. de Vries. Asset Market Linkages in Crisis Periods. *Review of Economics and Statistics*, 86(1):313–326, 2004.

P. Hartmann, S. Straetmans, and C.G. de Vries. Heavy Tails and Currency Crises. *Journal of Empirical Finance*, 17(2):241–254, 2010.

B.M. Hill. A Simple General Approach to Inference About the Tail of a Distribution. *Annals of Statistics*, 3(5):1163–1174, 1975.

J.B. Hill. On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays with Applications to Tail Index and Tail Dependence Estimation. *Journal of Statistical Planning and Inference*, 139(6):2091–2110, 2009.

J.B. Hill. Least Tail-trimmed Squares for Infinite Variance Autoregressions. *Journal of Time Series Analysis*, 34(2):168–186, 2013.

T. Hsing. On Tail Index Estimation Using Dependent Data. *Annals of Statistics*, 19(3): 1547–1569, 1991.

X. Huang. *Statistics of Bivariate Extreme Values*. PhD thesis, Erasmus University Rotterdam, 1992. Tinbergen Institute Research Series, PhD Thesis No. 22.

D.W. Jansen and C.G. De Vries. On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective. *Review of Economics and Statistics*, 73(1):18–24, 1991.

M.A. King and S. Wadhwani. Transmission of Volatility Between Stock Markets. *Review of Financial Studies*, 3(1):5–33, 1990.

R. Koenker and G. Bassett. Regression Quantiles. *Econometrica*, 46(1):33–50, 1978.

F. Longin and B. Solnik. Is the Correlation in International Equity Returns Constant: 1960-1990? *Journal of International Money and Finance*, 14(1):3–26, 1995.

F. Longin and B. Solnik. Extreme Correlation of International Equity Markets. *Journal of Finance*, 56(2):649–676, 2001.

Y. Malevergne and D. Sornette. How to Account for Extreme Co-Movements Between Individual Stocks and the Market. *Journal of Risk*, 6:71–116, 2004.

T. Mikosch and C.G. De Vries. Heavy Tails of OLS. *Journal of Econometrics*, 172(2): 205–221, 2013.

M. Mitchell and T. Pulvino. Characteristics of Risk and Return in Risk Arbitrage. *Journal of Finance*, 56(6):2135–2175, 2001.

L. Peng. Asymptotically Unbiased Estimators for the Extreme-Value Index. *Statistics & Probability Letters*, 38(2):107–115, 1998.

S.H. Poon, M. Rockinger, and J.A. Tawn. Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications. *Review of Financial Studies*, 17(2):581–610, 2004.

T. Post and P. Versijp. Multivariate Tests for Stochastic Dominance Efficiency of a Given Portfolio. *Journal of Financial and Quantitative Analysis*, 42(2):489–515, 2007.

K.D. Roose. The Recession of 1937-38. *Journal of Political Economy*, 56(3):239–248, 1948.

P.J. Rousseeuw. Multivariate Estimation with High Breakdown Point. In W. Grossman, G. Pflug, I. Vincze, and W. Wertz, editors, *Mathematical Statistics and Applications*, pages 283–297. Reidel Publishing Company, Dordrecht, 1985.

P. Sun and C. Zhou. Diagnosing the Distribution of GARCH Innovations. *Journal of Empirical Finance*, 29:287–303, 2014.

V. Todorov and T. Bollerslev. Jumps and Betas: A New Framework for Disentangling and Estimating Systematic Risks. *Journal of Econometrics*, 157(2):220–235, 2010.

M.R.C. Van Oordt and C. Zhou. Systematic Tail Risk. *Journal of Financial and Quantitative Analysis*, forthcoming.

W. Vervaat. Functional Central Limit Theorems for Processes with Positive Drift and Their Inverses. *Probability Theory and Related Fields*, 23(4):245–253, 1972.

Figure 1: Simulations with a global linear model

NOTE: The solid lines report the simulation results for the EVT approach; the dashed lines report those for the conditional regression approach. The simulations are based on 10,000 samples with 1,250 observations each. The observations of Y are constructed from the simulated observations of X and ε following different linear relations. The observations of X and ε are randomly drawn from the Student-t distribution with 4 degrees of freedom. The observations of Y are constructed from a global linear model ($\beta^T = \beta$). The estimates from the simulations, $\hat{\beta}^T$, are compared with the true value, β^T. The mean squared error is calculated as $\hat{\sigma}^2(\beta^T - \hat{\beta}_i^T)$, where i refers to the i-th simulated sample. The squared bias is calculated as $(\beta^T - \bar{\beta}^T)^2$ and the variance is calculated as $\hat{\sigma}^2(\bar{\beta}^T - \hat{\beta}_i^T)$, where $\bar{\beta}^T = \hat{\mu}(\hat{\beta}_i^T)$.

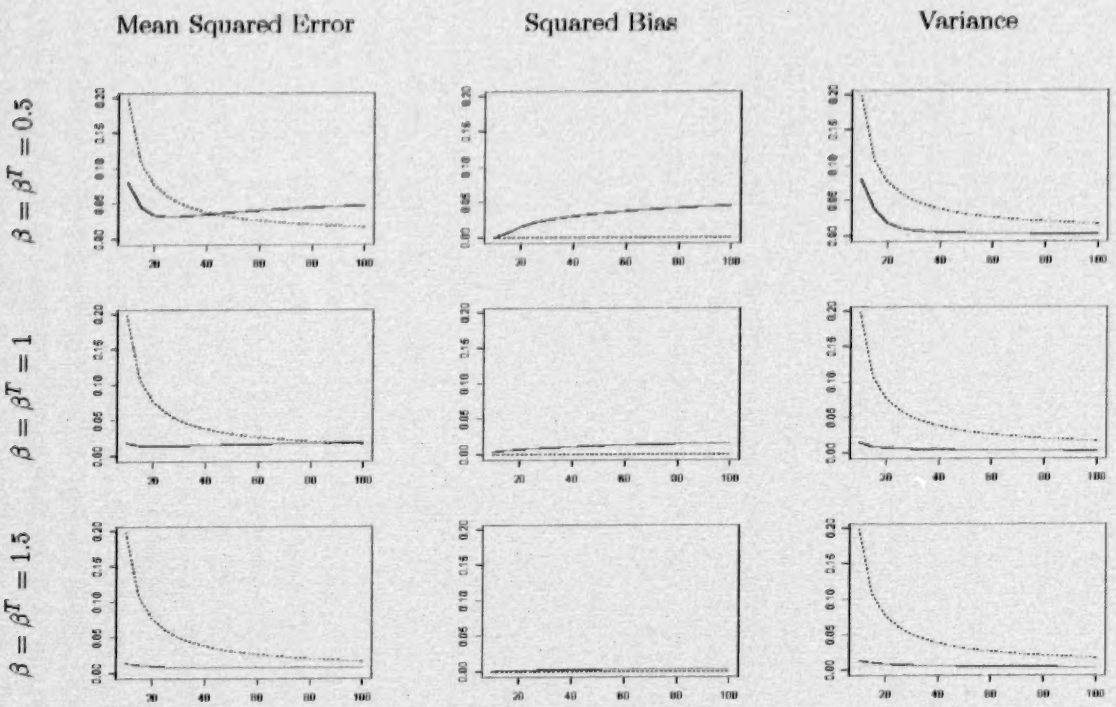

Mean Squared Error
Squared Bias
Variance
$\beta = \beta^T = 0.5$
$\beta = \beta^T = 1$
$\beta = \beta^T = 1.5$
Number of observations used in estimation, k

Figure 2: Simulations with a segmented linear model

NOTE: The solid lines report the simulation results for the EVT approach; the dashed lines report those for the conditional regression approach. The simulations are based on 10,000 samples with 1,250 observations each. The observations of Y are constructed from the simulated observations of X and ε following different linear relations. The observations of X and ε are randomly drawn from the Student-t distribution with 4 degrees of freedom. The observations of Y are constructed from a segmented linear model, where the slope equals β^T if the value of X_t is below its 3rd percentile (which occurs on expectation for 37.5 observations in each sample), and to β otherwise. The estimates from the simulations, $\hat{\beta}^T$, are compared with the true value, β^T. The mean squared error is calculated as $\hat{\sigma}^2(\beta^T - \hat{\beta}_i^T)$, where i refers to the i-th simulated sample. The squared bias is calculated as $(\beta^T - \bar{\beta}^T)^2$ and the variance is calculated as $\hat{\sigma}^2(\beta^T - \hat{\beta}_i^T)$, where $\bar{\beta}^T = \hat{\mu}(\hat{\beta}_i^T)$.

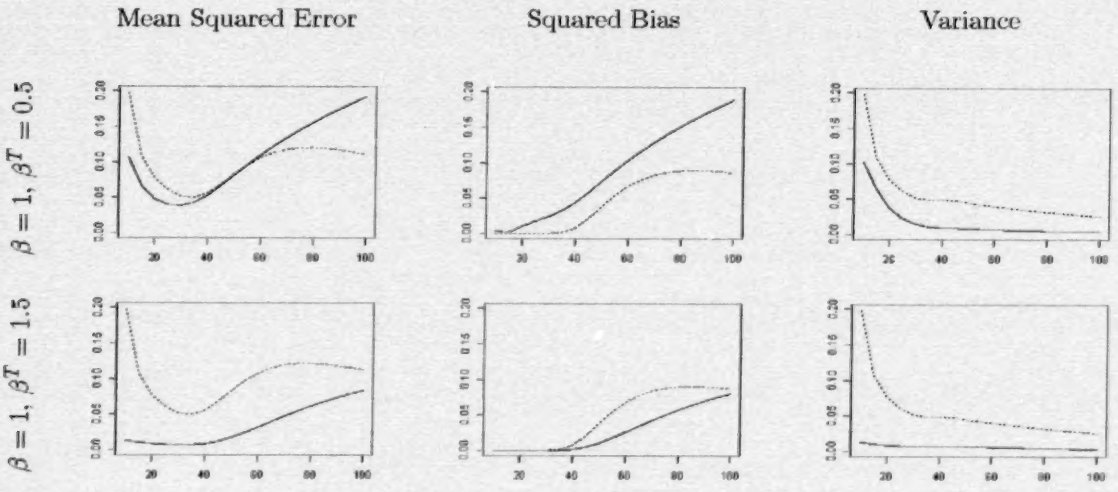


Table 1: Estimates

Period	Av. $\hat{\beta}^T_{EVT,j}$	N	S	Minimum $\hat{\beta}^T_{EVT,j}$		Maximum $\hat{\beta}^T_{EVT,j}$	
1931 - 1935	1.15	39	3	0.61	Tobacco Prdcts	1.79	Recreation
1936 - 1940	1.06	41	1	0.43	Tobacco Prdcts	2.08	Recreation
1941 - 1945	1.16	42	0	0.63	Communication	2.47	Real Estate
1946 - 1950	1.08	43	0	0.37	Communication	1.69	Construction
1951 - 1955	1.00	43	0	0.38	Communication	1.61	Aircraft
1956 - 1960	1.09	43	0	0.53	Food Products	1.71	Electronic Eq.
1961 - 1965	1.15	43	0	0.58	Utilities	1.93	Recreation
1966 - 1970	1.25	47	0	0.55	Utilities	1.90	Recreation
1971 - 1975	1.15	48	0	0.65	Utilities	1.78	Entertainment
1976 - 1980	1.09	45	3	0.61	Utilities	1.62	Healthcare
1981 - 1985	1.11	48	0	0.62	Utilities	2.31	Precious Metals
1986 - 1990	1.00	48	0	0.54	Utilities	1.32	Candy & Soda
1991 - 1995	1.16	48	0	0.64	Utilities	1.86	Shipbldng & Railrd Eq.
1996 - 2000	1.01	48	0	0.44	Utilities	1.86	Coal
2001 - 2005	1.02	48	0	0.56	Real Estate	1.70	Electronic Eq.
2006 - 2010	1.12	48	0	0.56	Beer & Liquor	2.19	Coal

NOTE: Within each period, we estimate coefficient β^T in (1.1) for the industry portfolios with non-missing observations using the EVT approach with $k = 25$, or, $k/n \approx 2\%$. For each period, we estimate β^Ts based on daily excess returns from that period while excluding the observation on the day the market suffered its largest loss. The column labeled 'Av. $\hat{\beta}^T_{EVT,j}$' reports the average of $\hat{\beta}^T_{EVT,j}$ of N portfolios. The column labeled 'S' reports the number of (excluded) portfolios with $\hat{\alpha}_j \leq \frac{1}{2}\hat{\alpha}_m$. The last columns report the minimum and maximum $\hat{\beta}^T_{EVT,j}$, and the industry name from the data documentation.

Table 2: Performance evaluation

Period	Worst day	Market loss	RMSE OLS	RMSE EVT	t-stat	p-value
1931 - 1935	21-Jul-1933	9.33	6.69	4.04	2.60	0.007
1936 - 1940	18-Oct-1937	8.10	4.93	2.59	3.18	0.001
1941 - 1945	08-Dec-1941	4.09	2.37	1.95	1.64	0.055
1946 - 1950	03-Sep-1946	6.82	2.39	1.45	1.68	0.050
1951 - 1955	26-Sep-1955	6.49	2.86	1.47	3.53	0.001
1956 - 1960	21-Oct-1957	3.04	1.96	1.20	2.98	0.002
1961 - 1965	28-May-1962	6.99	3.22	2.73	1.35	0.093
1966 - 1970	25-May-1970	3.21	2.23	1.42	2.63	0.006
1971 - 1975	18-Nov-1974	3.57	2.46	1.01	3.16	0.001
1976 - 1980	09-Oct-1979	3.44	2.07	0.85	5.42	0.000
1981 - 1985	25-Oct-1982	3.62	2.33	1.07	3.73	0.000
1986 - 1990	19-Oct-1987	17.44	6.73	3.95	1.34	0.094
1991 - 1995	15-Nov-1991	3.55	2.83	2.01	1.76	0.042
1996 - 2000	14-Apr-2000	6.73	4.35	3.37	1.11	0.136
2001 - 2005	17-Sep-2001	5.03	5.76	5.18	0.99	0.164
2006 - 2010	01-Dec-2008	8.95	3.37	1.40	3.59	0.000

NOTE: Within each period, we estimate the coefficient β^T in (1.1) for 48 industry portfolios using the EVT approach and the conditional regression approach with $k = 25$, or, $k/n \approx 2\%$. For each period, we estimate β^Ts based on daily excess returns during that period while excluding the observation on the day the market suffered its largest loss. Within each period, we project the loss on the day of the largest market loss for each industry portfolio j as the product of β_j^T and the actual market loss. From the difference between the projected losses on the industry portfolios and the actual losses, we calculate the root mean squared error (RMSE) for each approach. The last columns report the t-statistic and the corresponding probability for testing against the null hypothesis that the EVT approach produces a higher RMSE than the conditional regression (OLS) approach.